



احتمال

طراح: هادی بابالو

در یک مسابقه‌ی فوتبال ۴ تیم در مرحله‌ی گروهی حضور دارند. همه‌ی تیم‌ها دو به دو با هم بازی می‌کنند و هر دو تیم یک بازی با هم خواهند داشت. اگر بازی مساوی شود، دو تیم ۱ امتیاز خواهند گرفت و در غیر این صورت برنده ۳ امتیاز می‌گیرد و بازنده امتیازی نمی‌گیرد.

می‌دانیم احتمال برد تیم اول، برد تیم دوم و مساوی با هم برابر و مساوی $\frac{1}{3}$ است. تیم‌های a, b, c, d در یک گروه هستند و می‌دانیم در پایان مرحله گروهی تیم a توانسته ۶ امتیاز کسب کند. احتمال اینکه تیم b مرحله گروهی را با کسب ۴ امتیاز به اتمام رسانده باشد چقدر است؟

شبکه‌های بیزی

طراح: هادی بابالو

می‌خواهیم عملکرد دانشجویان یک کلاس را در امتحان درس هوش مصنوعی پیش‌بینی کنیم. می‌دانیم دانشجویانی عملکرد خوبی در امتحان دارند که به خوبی برای آن مطالعه کرده باشند و همین‌طور در زمان امتحان سردرد نداشته باشند. از طرفی در نظر می‌گیریم که سردرد یا ناشی از خستگی و یا ناشی از سینوزیت است. فرض کنید مطالعه، سینوزیت و خستگی دو به دو از هم **مستقل** هستند.

در نظر بگیرید که متغیر F نشان‌دهنده سینوزیت داشتن، T نشانه خسته بودن، H نشانه سردرد داشتن، S نشان‌دهنده مطالعه و در نهایت E نشان‌دهنده امتحان را خوب دادن باشد. توجه کنید که تمام متغیرهای تعریف شده **باینری** هستند.

الف) شبکه بیزی متناظر با این مسئله را رسم کنید.

ب) توزیع شبکه بیزی رسم شده را بنویسید.

ج) احتمال اینکه سنا سینوزیت داشته باشد، خسته نباشد، سردرد داشته باشد و امتحانش را خوب ندهد را بر اساس احتمال‌های شرطی و به ساده‌ترین شکل ممکن بنویسید.

د) فرض کنید که در این بخش می‌دانیم که خستگی روی مطالعه تاثیر می‌گذارد. با این فرض و در شبکه بیزی جدید درستی یا نادرستی موارد زیر را با ذکر دلیل **مختصر** مشخص کنید.

- متغیرهای F و T به شرط دانستن H از هم مستقل هستند.
- متغیرهای E و F به شرط دانستن H از هم مستقل هستند.
- متغیرهای F و T به طور کلی از هم مستقل هستند.
- متغیرهای S و H به طور کلی از هم مستقل هستند.

Hidden Markov Models

طراح: کیانوش عرشی، محمدطاها فخاریان

یک HMM با حالت‌های $\{a, b, c\}$ و الفبای $\{x, y, z\}$ تعریف کنید. احتمالات پایدار اولیه برای هر کدام از حالت‌ها برابر $\pi_a = 1$ ، $\pi_b = 0$ و $\pi_c = 0$ می‌باشند. همچنین احتمالات transition و emission در جدول زیر تعریف شده‌اند.

	a	b	c		x	y	z
a	0.2	0.8	0.0		0.8	0.2	0.0
b	0.0	0.8	0.2		0.0	0.6	0.4
c	0.4	0.0	0.6		0.2	0.0	0.8

- (الف) نمودار حالت¹ این HMM را رسم کنید و احتمالات transition را نشان دهید.
- (ب) تمام مسیرهای حالت با احتمال غیر صفر را برای دنباله $O = x, y, z, x$ ارائه دهید.
- (ج) $P(O)$ را به کمک brute force و Forward algorithm محاسبه کنید.
- (د) محتمل‌ترین مسیر (Q^*) چیست؟ احتمال $P(O, Q^*)$ (احتمال حرکت از این مسیر و نشر O) چیست؟ (می‌توانید از الگوریتم Viterbi استفاده کنید)
- (ه) برای یک دنباله مشخص O ، ممکن است احتمال $P(O)$ (احتمال اینکه مدل O را در تمامی راه‌های ممکن منتشر کند) به کمک $P(O, Q^*)$ (احتمال نشر O از طریق محتمل‌ترین مسیر) تخمین زده شود. برای HMM مشخص شده در این سوال، آیا مقدار $P(O, Q^*)$ تخمین خوبی برای $P(O)$ می‌باشد؟ پاسخ خود را شرح دهید.

Naive Bayes

طراح: اولدوز نیساری

به سوالات زیر در خصوص مفهوم naive bayes پاسخ تشریحی دهید.

- تفاوت بین diagnostic probability و causal probability را در naive bayes بیان کنید و با یک مثال توضیح دهید که چگونه naive bayes این دو را به یکدیگر مرتبط می‌کند؟
- در یک مطالعه پزشکی، متوجه شده‌اند که یک نوع بیماری تنها یک درصد از جامعه را تحت تاثیر قرار می‌دهد. یک تست تشخیصی انجام شده است. اگر فردی مبتلا باشد، این آزمایش به احتمال 99 درصد او را بیمار تشخیص می‌دهد، اما اگر فردی به بیماری مبتلا نباشد، به احتمال 5 درصد او را به اشتباه بیمار تشخیص می‌دهد. با توجه به داده‌های این مسئله توضیح دهید احتمال مثبت بودن نتیجه آزمایش در صورت ابتلا چه نقشی در محاسبه احتمال ابتلا در صورت مثبت بودن دارد؟ آیا وقتی از naive bayes برای استنتاج درباره این مسئله استفاده می‌کنیم، فرضی درباره استقلال پیشامدها داریم؟ احتمال بیمار نبودن را به شرط منفی بودن نتیجه تست را محاسبه کنید.

¹ State Diagram

Decision Trees

طراح: صادق فاضلی

آقای وحیدی که صاحب یک فروشگاه اینترنتی است، چندین کد تخفیف را از طریق پیامک برای تعدادی از مشتری‌های سابق فروشگاه ارسال کرده بود. بعضی از مشتریانی که کد تخفیف دریافت کرده بودند، از کد خود استفاده کرده و با خرید خود، آقای وحیدی را خوشحال کرده بودند. داده‌های مربوط مشتریان و استفاده آنها از کد تخفیف در جدول "داده‌های آموزش" قابل مشاهده است.

از آنجا که ارسال پیامک هزینه دارد، آقای وحیدی می‌خواهد بهینه عمل کرده و کدهای تخفیف جدید را فقط برای مشتریانی ارسال کند که انتظار می‌رود از فروشگاه خرید کنند. به همین دلیل آقای وحیدی از شما می‌خواهد یک Classifier طراحی کنید که بر اساس اطلاعات مشتری، استفاده کردن او از کد تخفیف را پیش‌بینی کند.

داده‌های آموزش					
	جنسیت	متاهل	سن	سابقه خرید در ماه گذشته	استفاده از تخفیف
۱	زن	بله	<۲۵	بله	✓
۲	مرد	خیر	۲۵-۴۰	بله	✓
۳	مرد	بله	۴۰<	خیر	×
۴	زن	خیر	<۲۵	بله	✓
۵	مرد	بله	۲۵-۴۰	خیر	×
۶	زن	بله	۲۵-۴۰	خیر	✓
۷	مرد	بله	۲۵-۴۰	بله	×
۸	مرد	بله	۲۵-۴۰	خیر	×
۹	زن	خیر	<۲۵	خیر	✓
۱۰	مرد	بله	۴۰<	خیر	✓
۱۱	زن	خیر	۲۵-۴۰	بله	✓
۱۲	مرد	خیر	۴۰<	بله	×
۱۳	زن	خیر	<۲۵	خیر	✓
۱۴	مرد	بله	<۲۵	بله	×
۱۵	زن	خیر	۴۰<	خیر	✓

قسمت اول

یک Classifier بر اساس Information Gain و با عمق ۳ (با احتساب ریشه و برگ‌ها) برای پیش‌بینی استفاده از کد تخفیف بسازید. از جدول "داده‌های آموزش" استفاده کنید. درخت نهایی و مراحل محاسبات خود را بنویسید.

قسمت دوم

با استفاده از Classifier ساخته شده در قسمت اول، "داده‌های آزمون" را پیش‌بینی کنید. سپس ماتریس درهم‌ریختگی (Confusion Matrix) را رسم کرده و Accuracy و Precision و Recall را محاسبه کنید.

داده‌های آزمون					
	جنسیت	متاهل	سن	سابقه خرید در ماه گذشته	استفاده از تخفیف
۱	مرد	بله	$40 <$	بله	✓
۲	زن	خیر	< 25	خیر	×
۳	مرد	بله	$25 - 40$	خیر	×
۴	زن	خیر	< 25	بله	✓
۵	مرد	خیر	$40 <$	بله	✓
۶	مرد	بله	$25 - 40$	خیر	×
۷	زن	خیر	< 25	خیر	✓
۸	زن	بله	< 25	بله	✓

قسمت سوم

به سوالات زیر به صورت تشریحی پاسخ دهید:

- 1) Classifier های درخت تصمیم چه زمانی دچار بیش‌برازش (Overfitting) می‌شوند؟ دلیل آن چیست؟
- 2) دو روش برای جلوگیری از بیش‌برازش (Overfitting) در Classifier های درخت تصمیم ارائه دهید.

Convolutional Neural Networks

طراح: علی محمدی

قسمت اول

شما در حال آموزش دادن یک Convolutional Neural Networks بر روی دیتاست ImageNet هستید، و شما در حال فکر کردن به این موضوع هستید که برای optimization function خود از gradient descent استفاده کنید. کدام گزینه/گزینه‌های زیر درست می‌باشد. توضیح کوتاه کافی است.

1. ممکن است که Stochastic Gradient Descent سریع‌تر از Batch Gradient Descent همگرا شود.
2. ممکن است که Mini Batch Gradient Descent سریع‌تر از Stochastic Gradient Descent همگرا شود.
3. ممکن است که Mini Batch Gradient Descent سریع‌تر از Batch Gradient Descent همگرا شود.
4. ممکن است که Batch Gradient Descent سریع‌تر از Stochastic Gradient Descent همگرا شود.

قسمت دوم

شما با یک تابع غیر خطی مواجه می‌شوید که اگر ورودی آن غیرمنفی باشد، عدد 1 را پاس می‌کند و در غیر اینصورت به 0 را می‌دهد، یعنی:

$$f(x) = \begin{cases} 1 & | \ x \geq 0 \\ 0 & | \ x < 0 \end{cases}$$

یکی از دوستان توصیه می‌کند که از این غیر خطی بودن در شبکه عصبی کانولوشنال خود با بهینه ساز Adam استفاده کنید. آیا به توصیه آنها عمل می‌کنید؟ چرا و چرا نه؟